1991

Comparative analysis of alternately spliced protein 4.1 mRNAs in erythroid and non-erythroid hematopoietic cells

Jean Louise Fraser
Yale University

Comparative Analysis of Alternately Spliced Protein 4.1 mRNAs in Erythroid and Non-Erythroid Hematopoietic Cells

Jean Louise Fraser

Yale University

1991

Permission for photocoping or microfilming of "Comparative Analysis of Alternately Spliced Protein 4.1 mRNAs in Erythroid and Non-Erythroid Hematopoietic Cells"
(Title of thesis)

for the purpose of individual scholarly consultation or reference is hereby granted by the author. This permission is not to be interpreted as affecting publication of this work or otherwise placing it in the public domain, and the author reserves all rights of ownership guaranteed under common law protection of unpublished manuscripts.

Glen L. Fraser
Signature of Author

5/17/91
Date

Comparative Analysis of Alternately Spliced Protein 4.1 mRNAs in Erythroid and Non-Erythroid Hematopoietic Cells

A Thesis Submitted to the Yale University School of Medicine in Partial Fulfillment of the Degree of Doctor of Medicine

by

Jean Louise Fraser

1991

ABSTRACT

COMPARATIVE ANALYSIS OF ALTERNATELY SPLICED PROTEIN 4.1 mRNAs IN ERYTHROID AND NON-ERYTHROID HEMATOPOIETIC CELLS

Jean Louise Fraser

1991

Protein 4.1 is a family of related protein isoforms generated by alternative mRNA splicing. The prototypical erythroid isoform is an 80 kD membrane skeleton component that serves to attach the cytoskeleton to the membrane through its binding interactions with transmembrane proteins and with other cytoskeletal elements. Non-erythroid isoforms, which localize to nuclear as well as cytoplasmic structures, differ from the erythroid isoform along eight nucleotide sequence motifs that are either inserted into or deleted from the primary mRNA transcript. I studied two RNA sequence motifs whose retention or omission affect both the size and apparent function of certain isoforms. The concerted splicing of these two motifs, IV and V, in non-erythroid cells provides a new in phase translation initiation site and allows manufacture of a 135 kD isoform. This larger isoform can be produced only if motif IV is retained and motif V is omitted from the final mRNA product. Using polymerase chain reaction (PCR) amplification and Southern blot analysis, I studied the pattern of these motif IV and V splicing events in erythroid

and non-erythroid cells. I found that the majority of transcripts in non-erythroid cells contain motif IV and lack motif V, resulting in generation of the large isoform. However, a minority of species encode shorter isoforms. This regulation of isoform production is probably accomplished via deletion of motif IV rather than insertion of motif V.

In a related set of experiments, I investigated the timing of splicing events during red cell maturation. Motif I is a nucleotide sequence block that is only inserted into transcripts of erythrocytes. I performed RNase protection assays and Northern blot analyses, but these failed to detect the presence of motif I in mouse erythroleukemia cells; subsequent studies by our group demonstrated induction of motif I between 86 and 110 hours after treatment with dimethyl sulfoxide, which is at least 48 hours after induction of the major erythrocyte protein, globin.

ACKNOWLEDGEMENTS

My sincerest gratitude is extended to Dr. Edward Benz, Jr. for his roles as advisor, mentor, and friend during the course of this work. His guiding influence provided direction and encouragement, yet allowed me the independence to develop my own ideas. The insight and experience I have gained is invaluable.

I wish also to thank Drs. Tang K. Tang and Lily Lou for their assistance and expertise at various stages of this project. Their advice was essential to the successful completion of several of my experiments. Finally, I want to express my appreciation to the other members of the Section of Hematology, especially Dr. Nancy Berliner, for their many helpful suggestions.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Protein 4.1 is a family of distinct but genetically related phosphoprotein isoforms exhibiting differential expression in both a tissue-specific and developmentally regulated manner. The prototypical isoform is a component of the erythrocyte membrane cytoskeleton.

The Erythrocyte Membrane Skeleton

The cytoskeleton is a filamentous intracellular protein network that serves a structural purpose in maintaining cell shape and plays a functional role in effecting cell movement, endocytosis, and surface receptor mobility. Most eucaryotic cells have a radial arrangement of actin filaments, intermediate filaments, and microtubules to accomplish this end. In erythrocytes, the corresponding structure is called the membrane skeleton and consists only of an array of interlocking proteins that underlies and supports the lipid bilayer (Figure 1). Red blood cells require an especially strong, yet flexible and deformable, membrane skeleton to withstand the shear stresses and many cell shape changes each undergoes during its transit through the circulation. A hexagonal protein lattice (Liu, 1987) periodically fixed to the overlying membrane is best adapted to meet these specialized needs.

Components

The erythrocyte membrane skeleton consists of a collection of large polypeptides originally identified as components I-VI by their relative positions on a sodium dodecyl sulfate polyacrylamide gel (Fairbanks, 1971). Further refinements of technique led to the discovery of additional protein species and gave rise to the now commonly used nomenclature for red blood cell membrane components (Steck, 1974) (Table 1).

Organization

The arrangement of major proteins in the erythrocyte membrane is represented schematically in Figure 2. In general, a meshwork of long, slender spectrin molecules cross-linked by nodular complexes of actin, protein 4.1, and protein 4.9 form the core of the membrane skeleton. This structure is then attached to the lipid bilayer via interactions with transmembrane proteins: ankyrin links spectrin heads to band 3 (an anion channel), and protein 4.1 connects spectrin tails to glycophorin (sialoglycoproteins bearing blood group antigens).

Role of Protein 4.1

Protein 4.1 participates in several of the key interactions of membrane skeleton components. The binding of protein 4.1 to spectrin greatly amplifies, strengthens, and stabilizes the association of spectrin tetramers and F-actin

Table 1. The major erythrocyte membrane proteins

Protein Band	Mol. Wt. (kD)	Other Names	Integral / Peripheral	Proportion (Wt %)	Copies / Cell Ghost	Chromosome Location
1	240	α-Spectrin	P	27	200,000	1q22-q25
2	220	β-Spectrin				14
2.1	210	Ankyrin	P	6	100,000	8p11-p21
2.2	195					
2.3	175					
2.6	145					
-	103	Adducin	P	1	30,000	
-	97					
3	100	Anion Channel	I	30	1,100,000	17
4.1	80/78		P	5	200,000	1p36.2-p34
4.2	72		P	5	250,000	
4.9	48		P	0.5	30,000	
5	43	Actin	P	5	400,000	7pter-q22
6	35	Glyceraldehyde-3-phosphate dehydrogenase	P	5	500,000	
7	29/27	Erythrocyte Tropomyosin	P	4	500,000	
8	23		P	1-2	200,000	
GPA	31	Glycophorin A	I	1.6	500,000	4q28-q31
GPB	23	Glycophorin B	I	0.3	100,000	4q28-q31
GPC	28	Glycophorin C	I	0.1	50,000	2q14-q21

Adapted from Lux, S.E., and Becker, P.S. Disorders of the Red Cell Membrane Skeleton. In The Metabolic Basis of Inherited Disease, 6th ed., Scriver, C.R., Beaudet, A.L., Sly, W.S., and Valle, D., eds.

(Ungewickell, 1979). This ternary complex imparts the necessary degree of rigidity to the cell membrane. Protein 4.1 also binds to the cytoplasmic domain of glycophorin (Anderson, 1984) in the presence of a specific polyphosphoinositide cofactor (Anderson, 1985). This interaction constitutes one class of attachment mechanisms for securing the cytoskeleton to the cell membrane. Finally, it has been observed that protein 4.1 can associate with band 3 as well (Pasternack, 1985). The use of this second, lower affinity transmembrane binding site may be related to the availability of phosphorylated metabolites.

Protein 4.1 in Abnormal Erythrocytes

The finding of quantitative or qualitative defects of protein 4.1 in clinically atypical erythrocytes confirms the role of this membrane skeleton constituent as a major determinant of red cell shape and durability. Studies of families with the condition hereditary elliptocytosis (HE) show that a subset of cases within this heterogeneous disorder are due to reduced levels of protein 4.1 (Tchernia, 1981). In the heterozygous form, a 50% decrease of protein 4.1 results in smooth, well-elongated elliptocytes with normal osmotic fragility. Clinically, little or no hemolysis occurs (Alloisio, 1985). Homozygotes, who are completely lacking protein 4.1, have a severe, transfusion-dependent hemolytic anemia with marked elliptocytosis and cell fragmentation (Tchernia, 1981). Normal mechanical stability can be restored

by addition of purified protein 4.1 to deficient cells (Takakuwa, 1986).

Another subset of HE cases can also be attributed to a protein 4.1 defect, despite the presence of normal protein 4.1 levels. Biochemical analysis of protein 4.1 polypeptides (Marchesi, 1990) and molecular analysis of mRNA transcripts (Conboy, 1990) reveals that insertion and deletion mutations produce structural variants of protein 4.1. The variants characterized to date have alterations in the spectrin-actin binding domain (see "Structural Model of Protein 4.1"). Deletion of the entire domain results in erythrocytes with markedly decreased mechanical stability, fragmented cells in the circulation, and a moderately severe hemolytic anemia. Duplication of the domain leads to only mild elliptocytosis without anemia.

Analogues of Protein 4.1 in Non-Erythroid Cells

Proteins immunologically cross-reactive with erythroid protein 4.1 are present in a wide variety of cell types. These species appear to function generally as linking molecules, although an exact parallel with the erythroid version is not uniformly applicable. In granulocytes (Spiegel, 1984) and platelets (Davies, 1985), the analogue is associated with the cytoskeleton, while in lymphoid cells (Spiegel, 1986) it is found in the cytoplasm. None of these proteins bind spectrin. An isoform in bovine aortic endothelial cells localizes to the cortical membrane, the

perinuclear region, and areas of cell-cell contact (Leto, 1986); the counterpart in fibroblasts is distributed evenly along stress fibers (Cohen, 1982). Protein 4.1-like proteins in these cells may serve to integrate actin filaments and other actin-binding skeletal components. The bovine lens also contains analogous forms which bind the lens version of spectrin and may play a role in the coupling of lens cytoskeleton to the plasma membrane.

The analogue of protein 4.1 immunologically isolated from neuronal cells is identical to a previously described protein named synapsin I (Baines, 1985). It has the same location as brain spectrin in the cortical cytoplasm (Goodman, 1984), and also binds to microtubules (Baines, 1986). Synapsin I is a known component of synaptic vesicles and is felt to have a linking function in joining vesicles to the plasma membrane or in attaching them to the axonal transport system.

Analogues of Protein 4.1 in Avian Erythrocytes

From extensive investigation of the avian erythrocyte system, a model of protein 4.1 expression has been developed which may provide interesting clues for analysis of the human system. The nucleated red cells of chickens contain several polypeptide variants, ranging in size from 77 kD to 175 kD (Granger, 1984). These isoforms are expressed in a developmentally regulated manner (Granger, 1985), with a transition from one set of forms predominately found in immature erythroblasts to another set of forms occurring

largely in mature erythrocytes (Yew, 1987). The finding of a single protein 4.1 gene that gives rise to numerous 6.6 kb mRNAs by differential RNA processing, provides a likely mechanism for the generation of these multiple, stage-dependent protein 4.1 variants (Ngai, 1987).

Structural Model of Protein 4.1

The prototypical erythroid protein 4.1 is actually a combination of two sequence-related phosphoproteins, designated 4.1a (80 kD) and 4.1b (78 kD) (Goodman, 1982). Limited proteolysis and specific chemical cleavage methods have enabled a detailed structural characterization of this protein (Leto, 1984) (Figure 3).

The 30 kD amino terminal domain is highly basic and resistant to further proteolytic digestion. Sites involved in binding protein 4.1 to cytoplasmic domains of the transmembrane proteins glycophorin and band 3 are located here (Leto, 1984). Sequence analysis shows this region to be hydrophobic and rich in beta-sheet structure. In addition, the 30 kD domain contains all seven cysteine residues and all three potential glycosylation sites present in the protein. These characteristics suggest a tightly compacted tertiary structure (Conboy, 1986a).

The 16 kD second domain (Leto, 1984) is hydrophilic and contains sites for phosphorylation by protein kinase C and cAMP-dependent kinase (Conboy, 1986a). This region has no known binding function and may simply act as a linker between

the membrane-glycophorin binding domain and the spectrin-actin binding domain.

The 10 kD third domain (Leto, 1984) is responsible for the binding of protein 4.1 to spectrin-actin complexes (Correas, 1986). It is highly charged and the secondary structure contains a relatively long alpha-helix. A site for cAMP-dependent kinase phosphorylation is also present (Conboy, 1986a).

The 22-24 kD carboxyl terminal domain is highly acidic and protease sensitive. The difference between the closely related proteins 4.1a and 4.1b lies in the post-translational modification of this domain (Leto, 1984).

Expression of the Protein 4.1 Gene

No cis- or trans-acting factors that influence transcription of the protein 4.1 gene have been identified. It is known that the synthesis of red cell membrane proteins is asynchronous during erythropoiesis and that protein 4.1 is among the last still being synthesized by reticulocytes (Chang, 1976). The relationship between expression of various components and assembly of the membrane skeleton is currently a topic of considerable interest. How the timing of expression of individual proteins might direct assembly of the cytoskeleton is uncertain.

Molecular Basis of Protein 4.1 Diversity

Cloning and sequencing of protein 4.1 cDNA (DNA

synthesized *in vitro* that is complementary to mRNA) has revealed that protein 4.1 is not a single entity but a heterogeneous family of related protein isoforms (Tang, 1988b). The isoforms are generated by alternative mRNA splicing of a common precursor and differ from one another at several specific sites. These nucleotide sequence segments, designated motifs I-VIII (Figure 4), are inserted or deleted *en bloc* and result in the production of numerous, distinct mRNAs. It now appears that multiple splicing pathways are operative even within a single cell type (Conboy, 1988), which then give rise to functionally divergent proteins. Aberrant splicing can lead to the manufacture of defective protein 4.1 species and to the clinical manifestations of HE described earlier (Conboy, 1986b).

Motif I encodes a 21 amino acid segment within the spectrin-actin binding domain. It is found in the protein 4.1 made by erythrocytes but not in that of lymphoid cells (Tang, 1988b). The ability of erythroid protein 4.1 to bind erythroid spectrin depends on the presence of the peptide piece determined by motif I (Horne, in press). Moreover, insertion of motif I is an inducible splicing event triggered during red cell maturation (L. Lou, unpublished data). Thus, motif I expression is governed by a tissue-specific and developmentally regulated splicing pathway.

Motifs II, III, VI, VII, and VIII are present in the vast majority of transcripts from both erythroid and non-erythroid cells (E. Benz, unpublished observations). Isoforms lacking

any of these motifs comprise only a fraction of the protein 4.1 in a given cell type, and the functional significance of these variants is unknown.

A Nuclear Protein 4.1 Isoform

Motifs IV and V are located 5' to the region of mRNA that is translated into the 80 kD protein 4.1 prototype by erythroid cells (Figure 4). Motif IV contains a new site for translation initiation, and motif V contains in phase stop codons. Two concerted splicing events, resulting in insertion of motif IV and deletion of motif V, place the new start site within a contiguous open reading frame that is in the same phase as the 80 kd protein (Tang; 1988a, 1990). Use of this upstream initiation codon allows synthesis of a larger 135 kD protein 4.1 isoform whose terminal 80 kD is identical to the prototypical erythroid isoform (Tang, 1990). When both start sites are present, the upstream codon is preferentially utilized (Tang; 1988a, 1990).

The high molecular weight variant of protein 4.1 includes an additional 209 amino acid polypeptide, corresponding to a calculated molecular weight of 27 kD, which is connected to the amino terminal end of the 30 kD domain of the erythroid protein 4.1 isoform. This headpiece is hydrophilic, contains three cysteine residues, and is rich in alpha-helical secondary structure (Tang, 1988a). Antibodies directed against the headpiece localize the 135 kD protein 4.1 isoform to the nuclei, as well as the cytoplasm, of non-erythroid

cells (Tang, in press). Current working hypotheses postulate a role for this nuclear isoform in linking the nuclear membrane to its skeleton or in integrating components of a nuclear scaffolding apparatus.

Thesis Goals

Recent studies (c.f. preceding sections) indicate that a single protein 4.1 gene gives rise to a family of protein 4.1 isoforms, some of which display tissue-specific expression. Regulation of isoform production appears to be at the level of mRNA splicing. Further understanding of isoform generation will require investigation of the mechanisms governing this splicing. However, it will first be necessary to characterize the patterns of motif expression in different tissues, with special attention to erythroid tissue, since tissue-specificity has already been demonstrated in mature red cells. Specific questions to be addressed include: Which motifs are developmentally induced or repressed, meaning spliced in or out in response to a given stimulus? Are motifs subject to individually regulated events or to concerted splicing, i.e. the splicing of one always accompanies splicing of the other? What is the timing of these splicing events during differentiation? An important focus will be the examination of the complex splicing events required for generation of the high molecular weight isoform. Production of this isoform is repressed during erythroid maturation but co-exists with production of low molecular weight variants in most other

tissues. For these reasons I chose to concentrate my initial efforts on analysis of motifs IV and V, as it is their splicing that determines which isoform, i.e. large or small, is manufactured.

The aim of this work was to investigate the expression of motifs IV and V in erythroid and non-erythroid cells. Nucleated cells produce both low and high molecular weight versions of protein 4.1. The splicing events required to generate the larger variant have been described; however, the smaller variant can result via a number of possible pathways: (1) deletion of motif IV, thereby removing the upstream initiation site; (2) insertion of motif V, which places stop codons in phase with the upstream start site and causes premature termination of translation; or (3) both deletion of motif IV and insertion of motif V. The research presented here examines the operative mechanism in elimination of the high molecular weight protein 4.1 isoform by non-erythroid, i.e. nucleated, cells.

Erythroid cells provide another setting in which to study the expression of motifs IV and V. During maturation from erythroblast to erythrocyte, the nuclei of these cells become pyknotic and are extruded. It is expected that production of the nuclear isoform of protein 4.1 will mirror this process. Accordingly, an inducible splicing event, such as that of motif I insertion, may also be governing expression of motif IV or V, or both. To that end, this research also examines the pattern of splicing of motifs IV and V during maturation

of an erythroid cell line.

Figure 1: The erythrocyte membrane skeleton contrasted with the cytoskeletal structures of most eucaryotic cells.

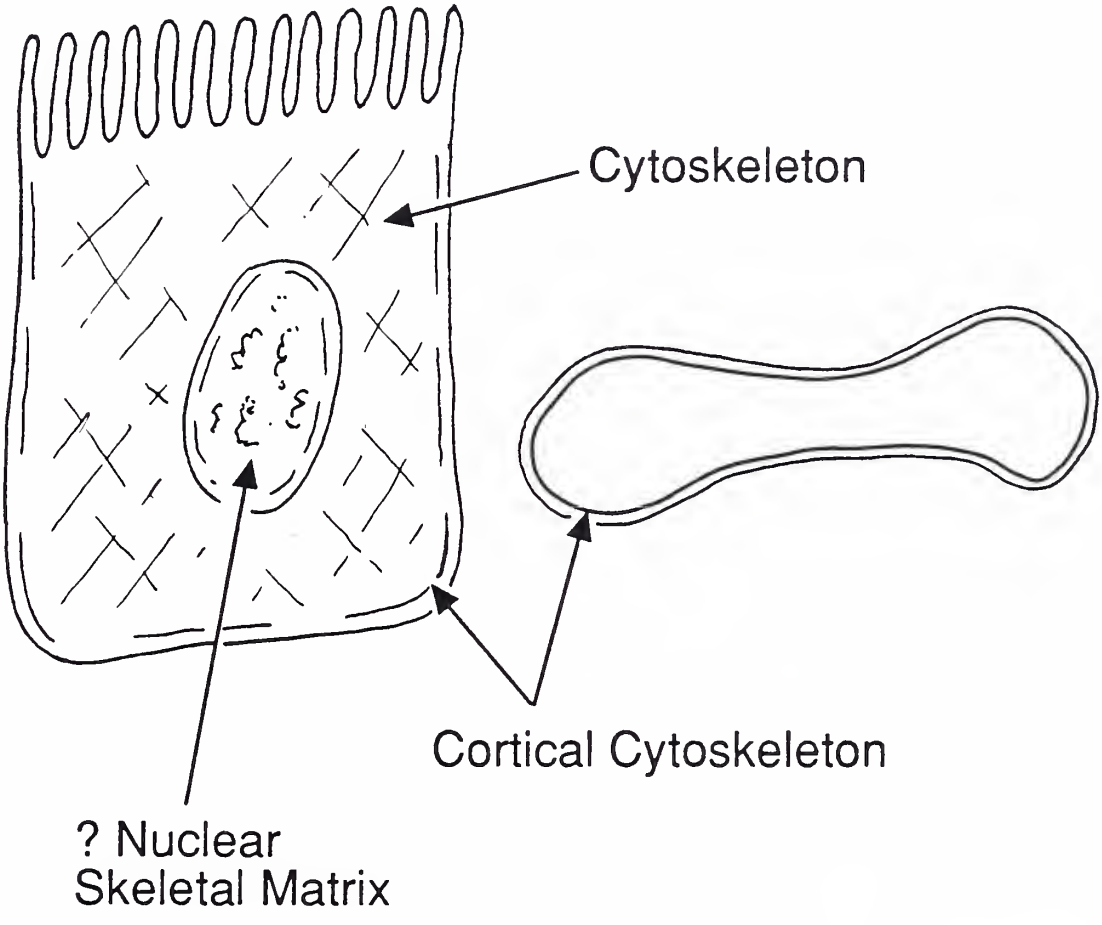


Figure 2: The erythrocyte membrane skeleton. Spectrin dimers composed of intertwined α- and β-chains self-associate at one end to form tetramers or higher oligomers (B) and bind to short filaments of F-actin and protein 4.1 at the opposite end (D). The membrane skeleton is attached to the lipid bilayer via interactions with ankyrin to the transmembrane protein band 3 (A) and with protein 4.1 to the transmembrane protein glycophorin (C) or protein 3. Tropomyosin (T) lies in the groove of the actin protofilament. (From Becker, P.S., and Benz, E.J., Jr. Molecular Biology of the Red Blood Cell Membrane Proteins. In Molecular Biology of the Cardiovascular System, Chien, S. editor. Copyright Lea and Febiger, Philadelphia, 1990. Used with permission.)

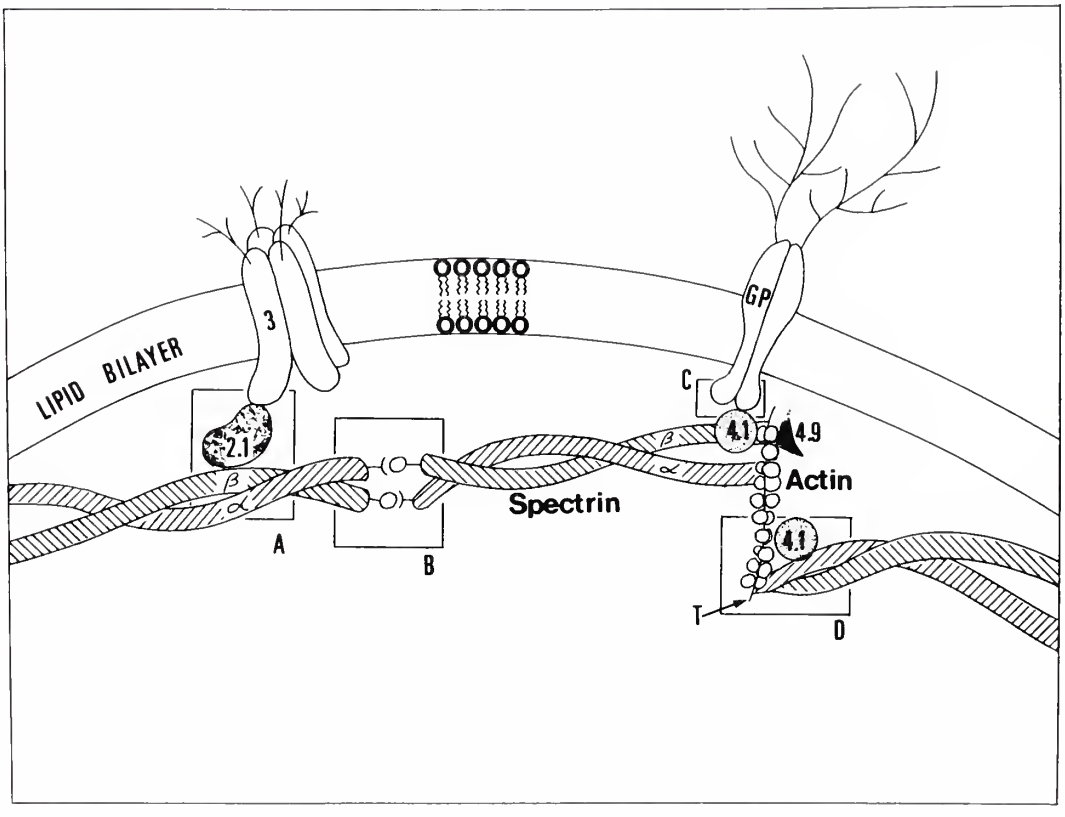


Figure 3: Structural model of protein 4.1.
A. The four domains of the protein 4.1 molecule identified by limited chymotryptic cleavage are diagrammed. Potential glycosylation sites are indicated by "G"; potential phosphorylation sites are indicated by "P"; "SH" denotes cysteine residues.
B. Secondary structural features of protein 4.1. Symbols are as follows:

 , α-helix;

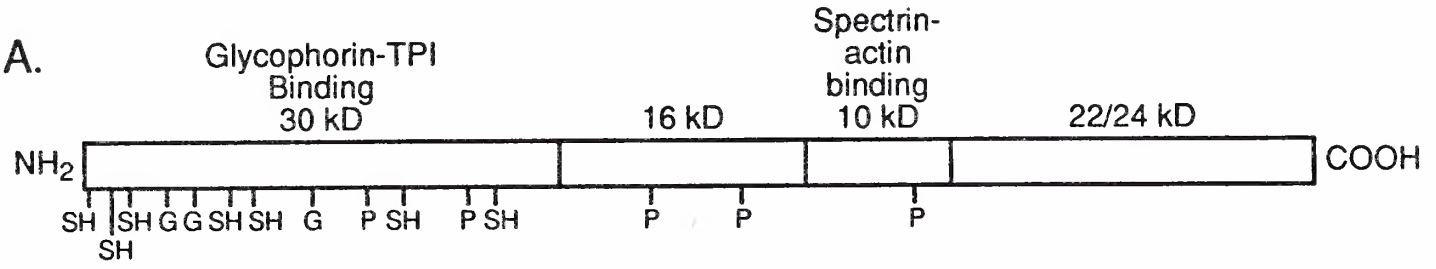
 , amphipathic α-helix;

 , hydrophobic α-helix;

 , β-sheet;

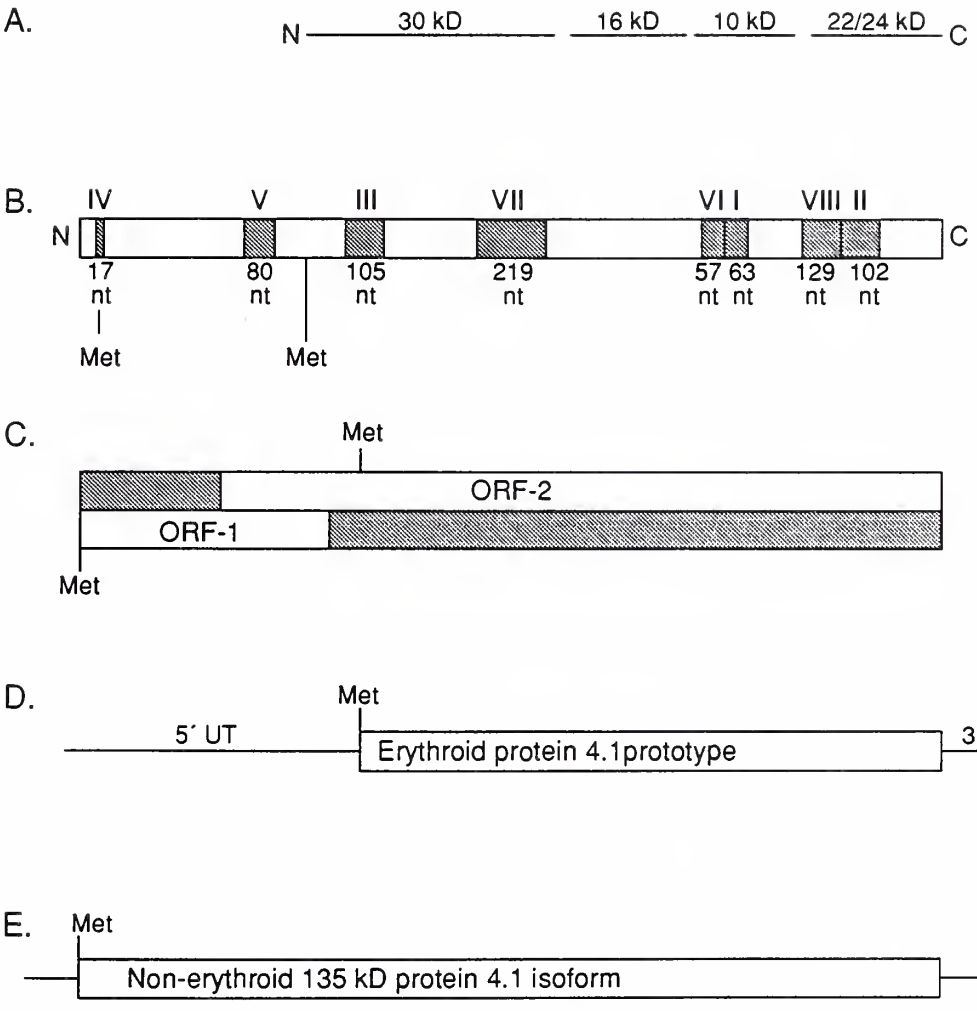
- - - -, random turns and coils.

(Adapted from Conboy, 1986a).


A.
Glycophorin-TPI
Binding
30 kD
16 kD
Spectrin-
actin
binding
10 kD
22/24 kD
NH2
COOH
SH
SH
SH
G
G
SH
SH
G
P
SH
P
SH
P
P
P


B.

Figure 4: Protein 4.1 isoforms.
A. Structural model of human erythroid protein 4.1. The four segments (30 kD, 16 kD, 10 kD, 22/24 kD) represent the molecular weights of the four fragments ("domains") generated by limited chymotryptic digestion. **B.** Location of motifs I - VIII. **C.** Two open reading frames (ORF 1 and 2) are found in the mRNA transcript. **D.** Region of mRNA translated to generate erythroid protein 4.1. **E.** Region of mRNA translated to generate non-erythroid 135 kD protein 4.1 isoform.


A.
N
30 kD
16 kD
10 kD
22/24 kD
C
B.
IV
V
III
VII
VI I
VIII II
N
C
17 nt
80 nt
105 nt
219 nt
57 63 nt nt
129 102 nt nt
Met
Met
C.
Met
ORF-2
ORF-1
Met
D.
Met
5´ UT
Erythroid protein 4.1prototype
3´ UT
E.
Met
Non-erythroid 135 kD protein 4.1 isoform

MATERIALS and METHODS

Cell Culture. Several human non-erythroid hematopoietic cell lines were studied. Previous work in this laboratory has shown that the available human erythroid cell lines do not differentiate completely enough to generate the erythroid isoforms of protein 4.1. Therefore, I focused my studies of erythroid isoforms on the mouse erythroleukemia (MEL) cell line. MEL cells are a permanent cell line from mice with a virally induced erythroleukemia. In the constitutive state, the cells are arrested at or near the proerythroblast stage of maturation. Upon exposure to any one of several inducing compounds (dimethyl sulfoxide, butyric acid, hexamethylene, bis-acetamide, etc.), the cells undergo erythroid maturation. The induced cells exhibit commitment to terminal maturation, as evidenced by *de novo* expression of numerous phenotypic markers of developing red cells (e.g. hemoglobin, erythroid spectrin, heme synthetic enzymes, etc.). These cells are the most thoroughly studied *in vitro* model of erythroid differentiation (Marks, 1978).

MEL cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum, 100 U/ml penicillin, 100 μg/ml streptomycin, and 2 mM L-glutamine. Induced cells were treated with 1.8% dimethyl sulfoxide (DMSO) and harvested 12 hr, 24 hr, and 48 hr after induction. Uninduced cell samples were also collected at each of the above time points to serve as controls.

My analysis of nonerythroid isoforms, especially the large isoform, focused on the following lymphoid and myeloid cell lines, which were known to produce the 135 kD isoform. These included HL60 (granulocytic/monocytic), KG-1 and KG-1a (monocytic), MOLT 4 (T-cell), CCRF-CEM and HUT 78 (B-cell). Cell lines were maintained in RPMI 1640 medium supplemented with 2 mM L-glutamine, 100 U/ml of penicillin, 100 μg/ml of streptomycin, and 10% fetal calf serum at 37°C in a 5% CO_2 atmosphere.

Total RNA Isolation. Total RNA from cultured cell lines was prepared in the presence of guanidine isothiocyanate (Davis, 1986). The cells (5 x 10^7) were pelleted by centrifugation at 2500 rpm in a HS-4 rotor for 5 minutes and resuspended in 8 ml of GIT buffer (4 M guanidine isothiocyanate / 25 mM sodium citrate / 7% β-mercaptoethanol). The GIT buffer lyses the cells, releasing nucleic acids, denaturing proteins, and increasing the viscosity of the solution.

The GIT-solubilized cells were layered onto 4 ml of CsCl buffer (5.7 M CsCl / 25 mM sodium acetate) and centrifuged at 32,000 rpm in a SW41 rotor for 18 hrs at 20°C. The clear gelatin-like RNA pellet at the bottom of the tube was collected and resuspended in 200 μl diethyl pyrocarbonate (DEPC) treated water. The RNA was further purified by extraction with saturated phenol/chloroform and precipitation with ethanol. The RNA was then recovered by centrifugation and dissolved in DEPC water.

Synthesis of High Specific Activity RNA Probes. Transcription plasmids containing cDNA inserts of interest and the T7 promoter were kindly provided by Dr. Tang K. Tang. Antisense RNA probes spanning motif I were generated by using T7 RNA polymerase. The following components were mixed in the order given in a 1.5 ml tube: 2.5 μl DEPC water, 4 μl 5X transcription buffer (200 mM Tris-HCl, pH 7.5 / 30 mM $MgCl_2$ / 10 mM spermidine / 50 mM NaCl), 2 μl 100 mM dithiothreitol, 40 U RNasin ribonuclease inhibitor, 4 μl rNTP (2.5 mM ATP / 2.5 mM GTP / 2.5 mM UTP), 2.4 μl 100 mM CTP, 1 μl containing 0.2-1 μg linearized DNA template, 5 μl α^{32}P-CTP (50 μCi), and 1 U T7 RNA polymerase. Incubation proceeded at 37°C for 60 minutes.

Following the RNA synthesis reaction, RNase free-DNase was added to a concentration of 1 U/μg DNA. Incubation was continued at 37°C for 15 minutes. The reaction mixture was extracted with an equal volume of phenol/chloroform. RNA was then recovered by precipitation with ethanol.

RNase Protection Assay. RNase protection analysis was performed using the uniformly labeled antisense RNA probes. Total RNA was hybridized to the antisense probe, followed by digestion with RNase A and T1.

Dried RNA samples (40 μg) were dissolved in 30 μl of formamide hybridization buffer (80% formamide / 40 mM PIPES, pH 6.5 / 0.4 M NaCl / 1 mM EDTA) containing antisense probe (5 x 10^5 cpm),heated at 85°C for 5 minutes, and then incubated at 45°C overnight. Following hybridization, 300 μl of RNase

digestion buffer (10 mM Tris-HCl, pH 7.5 / 5 mM EDTA / 300 mM NaCl) containing 40 μg of RNase A and 2 μg of RNase T1 was added. RNase digestion was allowed to proceed at 30°C for 1 hr.

The reaction was stopped by the addition of 20 μl of 10% SDS and 50 μg proteinase K, followed by incubation at 37°C for 15 minutes. Samples were phenol extracted, ethanol precipitated, dried, and dissolved in loading buffer (80% formamide, 0.025% bromophenol blue). The protected products were denatured for 3 minutes at 90°C and analyzed on a denaturing polyacrylamide / 8 M urea gel.

Poly (A)+ RNA Isolation. Poly (A)+ RNA was selected by hybridization to messenger affinity paper (mAP). 200 μg of total RNA was dissolved in 100 μl 0.5 M NaCl and spotted onto a 1 cm^2 piece of mAP. The mAP was then washed three times in 0.5 M NaCl and once in 70% ethanol. It was next blotted on filter paper and allowed to dry. The poly (A)+ RNA was eluted with 30 μl DEPC water in a 70°C water bath for 5 minutes.

End Labeling of Synthetic Oligonucleotide Probes. Synthetic oligonucleotides complementary to portions of motif I, motif IV, and motif V (Figure 5) were chemically synthesized and gel purified by Ms. Monica Talnor, Department of Pathology. These probes were labeled on the 5' ends using polynucleotide kinase (Davis, 1986). 0.25 μg of synthetic oligonucleotide was incubated with 15 U of T4 polynucleotide kinase and 20 μl of

γ^{32}P-ATP (200 μCi) in kinase buffer (50 mM Tris-HCl, pH 7.5 / 10 mM $MgCl_2$ / 5 mM dithiothreitol / 0.1 mM spermidine / 0.1 mM EDTA) for 1 hr at 37°C.

Removal of unincorporated nucleotides from the labeled DNA was accomplished by passing samples through activated Sep-Pak C18 cartridges. Columns were prewetted with 3 ml of 25% acetonitrile and 10 ml of distilled water. After application, the oligonucleotides were eluted with two separate 0.8 ml aliquots of 25% and 50% acetonitrile and then lyophilized. Probes were then resuspended in 200 μl of water; 1 μl was counted in a scintillation counter to determine the activity of the probes.

Northern (RNA) Blot Analysis. Poly (A)+ RNA was fractionated by electrophoresis through a 1.4% agarose formaldehyde gel (20 mM MOPS buffer, pH 7.0 / 5 mM sodium acetate / 0.5 mM EDTA / 2.2 M formaldehyde). The running buffer for the system was equivalent to the gel buffer without formaldehyde.

Following electrophoresis, the poly (A)+ RNA was transferred to Nytran filters and hybridized with the end labeled synthetic oligonucleotide probes. After baking the filters for 2 hr under vacuum at 80°C, prehybridization was performed in 6.7X SCP (1X SCP is 0.03 M $NaPO_4$, pH 6.5 / 0.1 M NaCl / 1 mM EDTA) / 10X Denhardt's solution (1X Denhardt's solution is 0.02% Ficoll 400 / 0.02% polyvinylpyrrolidone / 0.02% bovine serum albumin) / 1% sarkosyl / 1 μg/ml denatured salmon sperm DNA at 45°C for 1-2 hrs. The hybridization

solution contained the radiolabeled probe, carrier DNA (20 μg/ml final concentration), and 6.7X SCP / 10X Denhardt's solution / 1% sarkosyl / 10% dextran sulfate. The filters were incubated overnight at 45°C and then washed in 500 ml of 6.7X SCP / 0.5% sarkosyl, once at room temperature and three times at 45°C. After washing, the filters were baked and autoradiographed for up to 72 hrs.

Generation of cDNA from mRNA. Total RNA from the following cells was kindly provided by Dr. Lily Lou. Non-erythroid lines included HL60 (granulocytic/monocytic), KG-1 and KG-1a (monocytic), MOLT 4 (T-cell), CCRF-CEM and HUT 78 (B-cell); RNA from the erythroid MEL cell line had also been collected at 0, 27, 48, 60, 73, 86, 98, 110, 122, 167, and 186 hours after induction with DMSO. To prepare cDNA, 5 μg of total RNA was transcribed into single-stranded DNA at 42°C for 2 hrs in a 20 μl reaction containing 40 mM KCl / 50 mM Tris-HCl, pH 8.5 / 8 mM MgCl2 / 0.5 mM each dNTP / either 2 μg oligo dT primer, 4 μg dN_6 random hexamer primer, or 150 pmol specific antisense oligonucleotide / 10 U AMV reverse transcriptase / 40 U RNasin. Several primer strategies were attempted; the data presented in "Results" was obtained with oligo dT primer.

Primer Synthesis and Purification. Oligonucleotide primers for use in polymerase chain reaction (PCR) amplifications (Figure 6) were chemically synthesized by Ms. Monica Talnor, Department of Pathology (motifs I and V), and Dr. John Flory,

Department of Human Genetics (motif IV). Primers received in an unpurified form were passed through an activated Sep-Pak C18 cartridge as described in a previous section (see "End Labeling of Synthetic Oligonucleotide Probes").

PCR Amplification of Motifs I, IV, and V. Amplification of motif I in MEL cells was performed by Dr. Lily Lou. The PCR amplification of motifs IV and V in non-erythroid and MEL cells contained 2 μl cDNA / 10 pmol sense-strand primer / 10 pmol antisense-strand primer / 0.2 mM each dNTP / 2.5 U Taq polymerase in PCR buffer (10 mM Tris-HCl, pH 8.5 / 50 mM KCl / 1.5 mM $MgCl_2$ / 0.1% gelatin). Thirty cycles of amplification were performed using an automated Perkin-Elmer Cetus thermal cycler under the following conditions: melting, 1.5 min at 94°C; annealing, 2 min at 55°C; polymerizing, 3 min at 72°C.

Southern (DNA) Blot Analysis. PCR products were analyzed by 2% agarose gel electrophoresis in TBE buffer (50 mM Tris-HCl, pH 7.5 / 50 mM boric acid / 1 mM EDTA). After electrophoresis, gels were washed in 200 ml of denaturing solution (1.5 M NaCl / 0.5 N NaOH) followed by 200 ml of neutralizing solution (1.5 M NaCl / 0.5 M Tris-HCl, pH 7.5) for 1 hr each. DNA fragments were then transferred to nitrocellulose filters and hybridized with end labeled synthetic oligonucleotide probes to the motif of interest. Conditions for oligonucleotide hybridization and washing were

as previously described (see "Northern Blot Analysis") with slight modifications. Baked filters were prehybridized in 6X SSC (1X SSC is 0.15 M NaCl / 15 mM Na citrate) / 5X Denhardt's solution / 0.5% SDS / 1% denatured salmon sperm DNA at room temperature for 3 hrs. The hybridization solution was identical to the prehybridization buffer with addition of 0.01 M EDTA and the radiolabeled probe. Hybridization proceeded overnight at 31°C for motif IV and at 53°C for motif V. The filters were then washed four times: 5 min in 2X SSC / 0.5% SDS, 15 min in 2X SSC / 0.1% SDS, 15 min in 0.4X SSC / 0.1% SDS, and 15 min in 0.1X SSC / 0.1% SDS. After washing, the filters were baked and autoradiographed for 48 hours.

Figure 5: Synthetic oligonucleotide probes used to identify motifs I, IV, and V in Northern and Southern blot analyses. **A.** Motif I. **B.** Motif IV and motif V.

A.

```
   1  TGGCCCAAAG TGCTGAAGAT TTCTTATAAA CGTAGTAGCT TTTTCATCAA GATTCGGCCT
  61  GGAGAGCAAG AGCAGTATGA AAGTACCATC GGATTCAAAC TTCCCAGTTA CCGAGCAGCT
 121  AAGAAATTAT GGAAAGTCTG TGTAGAACAT CACACGTTTT TCAGATTGAC ATCTACAGAC
 181  ACCATTCCCA AAAGCAAATT TCTTGCGCTA GGATCCAAAT TTCGATACAG TGGCCGGACT
 241  CAAGCTCAGA CCAGGCAAGC TAGTGCTCTA ATTGACAGGC CTGCCCCACA CTTCGAGCGT
 301  ACAGCAAGTA AACGGGCGTC CCGGAGCCTC GATGGAGCAG CAGCTGTCGA TTCGGCAGAC
 361  CGAAGTCCTC GGCCCACTTC TGCACCTGCC ATTACTCAGG GTCAGGTTGC AGAAGGTGGC
 421  GTCCTAGATG CCTCTGCTAA AAAAACAGTG GTCCCTAAAG CACAGAAGGA AACAGTGAAG
 481  GCTGAAGTGA AAAAGGAAGA CGAGCCACCT GAGCAAGCTG AGCCAGAGCC CACAGAAGCA

                                        Motif I
 541  TGGAAGAAAA AGAGAGAAAG ACTAGATGGT GAAAACATTT ATATCAGACA TAGCAATTTA
                T TCTCTCTTTC TGATCTACCA CTTTTGTAAA TATAGTCTG

 601  ATGTTGGAGG ATTTAGACAA GAGTCAAGAG GAGATCAAAA AACATCATGC CAGCATCAGT
 661  GAGCTGAAAA AGAACTTCAT GGAGTCTGTA CCAGAACCAC GGCCTAGTGA ATGGGATAAA
 721  CGCTTATCCA CTCACTCACC CTTCCGAACT CTTAACATCA ATGGGCAAAT CCCCACAGGA
 781  GAAGGACCTC CCCTGGTGAA GACACAAACT GTCACCATCT CAGATAATGC CAATGCTGTG
 841  AAAAGTGAAA TCCCAACCAA AGACGTCCCT ATTGTCCACA CTGAGACCAA GACCATCACT
 901  TATGAGGCTG CCCAG
```

B.

```
   1  TCCCGTGGAG CAGAGGGGCA AAGTGGCAGG AACCTCTTAA AGGGCGAGAG CGGCGCGGAG

                                                          Motif IV
  61  CCAGAACGCG GTCGGCCCGG TCCCCGCCGC ACCCAGCCCA GCAACATCAT GACAACAGAG
                                                       GTTGTAGTA CTGTTGTC

 121  AAGAGTTTAG TGACTGAGGC CGAAAATTCA CAGCACCAAC AGAAGGAAGA GGGTGAGGAA

 181  GCCATAAACT CAGGCCAACA AGAACCTCAG CAGGAGGAAT CTTGTCAAAC AGCAGCTGAA

 241  GGAGATAATT GGTGTGAACA GAAGCTGAAA GCTTCTAATG GAGACACTCC TACACATGAA

 301  GACTTGACCA AGAACAAGGA GCGGACATCA GAAAGCAGAG GACTTTCACG ACTATTCTCC

 361  TCGTTTCTCA AAAGGCCCAA ATCTCAGGTG TCCGAGGAAG AAGGCAAAGA AGTAGAGTCA

 421  GATAAAGAAA AAGGTGAAGG AGGTCAGAAA GAGATAGAAT TTGGAACCAG TCTTGATGAA

 481  GAGATCATTT TAAAGGCCCC AATTGCAGCT CCTGAACCGG AACTCAAAAC AGACCCATCT

 541  TTGGATCTTC ATTCATTAAG CAGTGCAGAA ACACAGCCTC ACCATTACAA TTAAGAATTA
                                                                    TAAT

                                      Motif V
 601  TTTTTAGAGT CTTCTTATTT CTGAAGCATG TGAATATTAT CTCGATCGTT AAAAGTCCTG
      AAAAATCTCA GAAGAATAAA GACTTCGTAC ACTTATAATA GAGCTAGCAA

 661  CTCAGGAAGA ACTCAGAGAA GATCCAGATT TTGAAATTAA GGAAGGAGAA GGACTTGAAG

 721  AGTGCTCCAA AATAGAAGTA AAAGAAGAAA GCCCTCAATC AAAAGCAGAA ACAGAATTAA

 781  AAGCTTCCCA AAAACCAATC AGAAAACACA GGAACATGCA CTGCAAGGTT TCTTTGTTGG

 841  ATGACACAGT TTATGAATGT GTTGTGGAG
```

Figure 6: Synthetic oligonucleotide primers used in PCR amplification of motifs I, IV, and V. **A.** Motif I. **B.** Motif IV and motif V.

A.

```
   1  TGGCCCAAAG TGCTGAAGAT TTCTTATAAA CGTAGTAGCT TTTTCATCAA GATTCGGCCT
      ACCGGGTTTC ACGACTTCTA AAGAATATTT GCATCATCGA AAAAGTAGTT CTAAGCCGGA

  61  GGAGAGCAAG AGCAGTATGA AAGTACCATC GGATTCAAAC TTCCCAGTTA CCGAGCAGCT
      CCTCTCGTTC TCGTCATACT TTCATGGTAG CCTAAGTTTG AAGGGTCAAT GGCTCGTCGA

 121  AAGAAATTAT GGAAAGTCTG TGTAGAACAT CACACGTTTT TCAGATTGAC ATCTACAGAC
      TTCTTTAATA CCTTTCAGAC ACATCTTGTA GTGTGCAAAA AGTCTAACTG TAGATGTCTG

 181  ACCATTCCCA AAAGCAAATT TCTTGCGCTA GGATCCAAAT TTCGATACAG TGGCCGGACT
      TGGTAAGGGT TTTCGTTTAA AGAACGCGAT CCTAGGTTTA AAGCTATGTC ACCGGCCTGA

 241  CAAGCTCAGA CCAGGCAAGC TAGTGCTCTA ATTGACAGGC CTGCCCCACA CTTCGAGCGT
      GTTCGAGTCT GGTCCGTTCG ATCACGAGAT TAACTGTCCG GACGGGGTGT GAAGCTCGCA

                                                        →
 301  ACAGCAAGTA AACGGGCGTC CCGGAGCCTC GATGGAGCAG CAGCTGTCGA TTCGGCAGAC
      TGTCGTTCAT TTGCCCGCAG GGCCTCGGAG CTACCTCGTC GTCGACAGCT AAGCCGTCTG

 361  CGAAGTCCTC GGCCCACTTC TGCACCTGCC ATTACTCAGG GTCAGGTTGC AGAAGGTGGC
      GCTTCAGGAG CCGGGTGAAG ACGTGGACGG TAATGAGTCC CAGTCCAACG TCTTCCACCG

 421  GTCCTAGATG CCTCTGCTAA AAAAACAGTG GTCCCTAAAG CACAGAAGGA AACAGTGAAG
      CAGGATCTAC GGAGACGATT TTTTTGTCAC CAGGGATTTC GTGTCTTCCT TTGTCACTTC

 481  GCTGAAGTGA AAAAGGAAGA CGAGCCACCT GAGCAAGCTG AGCCAGAGCC CACAGAAGCA
      CGACTTCACT TTTTCCTTCT GCTCGGTGGA CTCGTTCGAC TCGGTCTCGG GTGTCTTCGT

                                        Motif I
 541  TGGAAGAAAA AGAGAGAAAG ACTAGATGGT GAAAACATTT ATATCAGACA TAGCAATTTA
      ACCTTCTTTT TCTCTCTTTC TGATCTACCA CTTTTGTAAA TATAGTCTGT ATCGTTAAAT

 601  ATGTTGGAGG ATTTAGACAA GAGTCAAGAG GAGATCAAAA AACATCATGC CAGCATCAGT
      TACAACCTCC TAAATCTGTT CTCAGTTCTC CTCTAGTTTT TTGTAGTACG GTCGTAGTCA

 661  GAGCTGAAAA AGAACTTCAT GGAGTCTGTA CCAGAACCAC GGCCTAGTGA ATGGGATAAA
      CTCGACTTTT TCTTGAAGTA CCTCAGACAT GGTCTTGGTG CCGGATCACT TACCCTATTT

 721  CGCTTATCCA CTCACTCACC CTTCCGAACT CTTAACATCA ATGGGCAAAT CCCCACAGGA
      GCGAATAGGT GAGTGAGTGG GAAGGCTTGA GAATTGTAGT TACCCGTTTA GGGGTGTCCT
                                                                    ←

 781  GAAGGACCTC CCCTGGTGAA GACACAAACT GTCACCATCT CAGATAATGC CAATGCTGTG
      CTTCCTGGAG GGGACCACTT CTGTGTTTGA CAGTGGTAGA GTCTATTACG GTTACGACAC

 841  AAAAGTGAAA TCCCAACCAA AGACGTCCCT ATTGTCCACA CTGAGACCAA GACCATCACT
      TTTTCACTTT AGGGTTGGTT TCTGCAGGGA TAACAGGTGT GACTCTGGTT CTGGTAGTGA

 901  TATGAGGCTG CCCAG
      ATACTCCGAC GGGTC
```

B.

```
   1  TCCCGTGGAG CAGAGGGGCA AAGTGGCAGG AACCTCTTAA AGGGCGAGAG CGGCGCGGAG
      AGGGCACCTC GTCTCCCCGT TTCACCGTCC TTGGAGAATT TCCCGCTCTC GCCGCGCCTC

                                                              Motif IV
  61  CCAGAACGCG GTCGGCCCGG TCCCCGCCGC ACCCAGCCCA GCAACATCAT GACAACAGAG
      GGTCTTGCGC CAGCCGGGCC AGGGGCGGCG TGGGTCGGGT CGTTGTAGTA CTGTTGTCTC

 121  AAGAGTTTAG TGACTGAGGC CGAAAATTCA CAGCACCAAC AGAAGGAAGA GGGTGAGGAA
      TTCTCAAATC ACTGACTCCG GCTTTTAAGT GTCGTGGTTG TCTTCCTTCT CCCACTCCTT

 181  GCCATAAACT CAGGCCAACA AGAACCTCAG CAGGAGGAAT CTTGTCAAAC AGCAGCTGAA
      CGGTATTTGA GTCCGGTTGT TCTTGGAGTC GTCCTCCTTA GAACAGTTTG TCGTCGACTT

 241  GGAGATAATT GGTGTGAACA GAAGCTGAAA GCTTCTAATG GAGACACTCC TACACATGAA
      CCTCTATTAA CCACACTTGT CTTCGACTTT CGAAGATTAC CTCTGTGAGG ATGTGTACTT

 301  GACTTGACCA AGAACAAGGA GCGGACATCA GAAAGCAGAG GACTTTCACG ACTATTCTCC
      CTGAACTGGT TCTTGTTCCT CGCCTGTAGT CTTTCGTCTC CTGAAAGTGC TGATAAGAGG

 361  TCGTTTCTCA AAAGGCCCAA ATCTCAGGTG TCCGAGGAAG AAGGCAAAGA AGTAGAGTCA
      AGCAAAGAGT TTTCCGGGTT TAGAGTCCAC AGGCTCCTTC TTCCGTTTCT TCATCTCAGT

 421  GATAAAGAAA AAGGTGAAGG AGGTCAGAAA GAGATAGAAT TTGGAACCAG TCTTGATGAA
      CTATTTCTTT TTCCACTTCC TCCAGTCTTT CTCTATCTTA AACCTTGGTC AGAACTACTT

 481  GAGATCATTT TAAAGGCCCC AATTGCAGCT CCTGAACCGG AACTCAAAAC AGACCCATCT
      CTCTAGTAAA ATTTCCGGGG TTAACGTCGA GGACTTGGCC TTGAGTTTTG TCTGGGTAGA

 541  TTGGATCTTC ATTCATTAAG CAGTGCAGAA ACACAGCCTC ACCATTACAA TTAAGAATTA
      AACCTAGAAG TAAGTAATTC GTCACGTCTT TGTGTCGGAG TGGTAATGTT AATTCTTAAT

                                  Motif V
 601  TTTTTAGAGT CTTCTTATTT CTGAAGCATG TGAATATTAT CTCGATCGTT AAAAGTCCTG
      AAAAATCTCA GAAGAATAAA GACTTCGTAC ACTTATAATA GAGCTAGCAA TTTTCAGGAC

 661  CTCAGGAAGA ACTCAGAGAA GATCCAGATT TTGAAATTAA GGAAGGAGAA GGACTTGAAG
      GAGTCCTTCT TGAGTCTCTT CTAGGTCTAA AACTTTAATT CCTTCCTCTT CCTGAACTTC

 721  AGTGCTCCAA AATAGAAGTA AAAGAAGAAA GCCCTCAATC AAAAGCAGAA ACAGAATTAA
      TCACGAGGTT TTATCTTCAT TTTCTTCTTT CGGGAGTTAG TTTTCGTCTT TGTCTTAATT

 781  AAGCTTCCCA AAAACCAATC AGAAAACACA GGAACATGCA CTGCAAGGTT TCTTTGTTGG
      TTCGAAGGGT TTTTGGTTAG TCTTTTGTGT CCTTGTACGT GACGTTCCAA AGAAACAACC

 841  ATGACACAGT TTATGAATGT GTTGTGGAG
      TACTGTGTCA AATACTTACA CAACACCTC
```

RESULTS and DISCUSSION

Expression of Motif I during Erythroid Cell Maturation. In order to assess the pattern of motif I expression in mRNA of erythroid cells during maturation, I performed RNase protection assays and Northern blot analyses on RNA samples collected from MEL cells at 0, 24, and 48 hrs after induction with DMSO. An RNase protection assay allows a specific comparison between a cell's RNA and a labeled antisense RNA probe. If the probe does not base pair precisely to the RNA, extra tails or loops not protected by RNA hybridization will be excised by RNase digestion. Therefore, probes designed to span motif I can generate distinct protected fragments when hybridized to mRNAs with or without motif I.

Northern blot analysis also identifies homologous RNA species in a cell's mRNA. This method detects sequences complementary to a labeled antisense probe via hybridization with poly (A)+ RNA bound to a filter. These probes are derived from sequences present within motif I, and thus will bind only to RNA that contains motif I.

Both of these studies produced similar results in that neither showed specific hybridization to the antisense motif I probes. Possible explanations include lack of homology between mouse (RNA) and human (probes) sequences or inadequate amounts of RNA containing motif I, especially in the early stages of induction. The RNase protection assay is quite sensitive but is susceptible to excessive probe digestion if

the probe and RNA are insufficiently homologous. The Northern blot is capable of detecting modestly homologous RNAs but is not especially sensitive. Therefore, the most likely explanation of these disappointing results is that the RNase protection assay failed to work because of inadequate homology; the Northern blot assay is more tolerant of partial homology but is far less sensitive.

On the basis of these findings, Dr. Lily Lou subsequently elucidated the pattern of motif I expression during erythroid cell maturation using PCR. This technique specifically replicates a particular segment of DNA by flanking the fragment with oligonucleotide primers and repeatedly denaturing, annealing, and extending the primers. These primers hybridize to opposite strands of the target sequence and are oriented so that DNA synthesis by a heat-stable polymerase proceeds across the region between the primers. Since the extension products themselves are also complementary to and capable of binding primers, successive cycles of amplification result in an exponential accumulation of the specific target fragment. By amplifying the region of mRNA containing motif I, Dr. Lou showed that barely detectable amounts of motif I are present in the uninduced cells. An appreciable increase in motif I expression occurred between 86 and 110 hours, and a decline in expression began after 122 hours of induction. The induction of motif I takes place at least 48 hours after globin induction (Figure 7).

Expression of Motifs IV and V in Non-Erythroid Hematopoietic Cells. PCR amplifications of sequences spanning motif IV and motif V were targeted by using 5' and 3' flanking oligonucleotides for each motif. The expected products obtained by amplifying cDNA templates that include or exclude motif IV were 96 bp and 79 bp, respectively and 258 bp and 178 bp, respectively for motif V. Use of the 5' motif IV primer and the 3' motif V primer allows analysis of both motifs as they occur on a single template; the expected products in this case are 626 bp (+IV, +V), 546 bp (+IV, -V), 609 bp (-IV, +V), and 529 bp (-IV, -V). Amplification of cDNA generated from the non-erythroid cell lines HL60, KG-1, KG-1a, MOLT 4, CCRF-CEM, and HUT 78 was performed.

Analysis of the PCR products by ethidium bromide staining of agarose gels generally reveals only the most prominent products, as shown in Figure 8. In this case, the vast majority of RNA species in non-erythroid cells exhibit only a single band, 178 bp in length, when amplified with primers flanking motif V. These results suggest that the bulk of protein 4.1 mRNA present in the cells lacks motif V. When analyzed with primers flanking motif IV, these RNAs generated more complex results (Figure 9). Two bands, 96 bp and 79 bp, of roughly equal intensity are seen in most of the cell lines. These bands correspond well to the predicted sizes for RNAs containing and lacking motif IV. Analysis with primers yielding a PCR product that spans both motifs (Figure 10) showed only a single band, 546 bp, consistent with absence of

motif V and presence or absence of motif IV. (Note that the 17 bp difference between products containing and lacking motif IV cannot be resolved on this gel due to the large size of the fragments).

The PCR products were further analyzed by Southern blot analysis using oligonucleotides complementary to each motif (Figure 5B). This technique will detect minor species as well as the more predominant products detected by ethidium bromide staining. However, neither method can be utilized for quantitative assessment of relative amounts of individual motifs. When the products amplified by primers flanking motif V were hybridized to a probe complementary to sequences within motif V, products containing motif V could then be visualized (Figure 11). These migrated to a position corresponding to a size roughly 80 bp larger than the ethidium bromide stained bands, which, of course, did not hybridize to the probe because they lack motif V. Similarly, Southern blot analysis of products generated by motif IV amplification (Figure 12) showed a predominant band at the size expected for retention of motif IV as well as some minor bands of higher molecular weight, the identity of which remain unknown. Finally, Figure 13 shows that at least a portion of the products spanning both motifs contain motif IV. This experiment suggests that at least some transcripts lacking motif V contain motif IV, but that no transcripts containing motif V also contain motif IV. Confirmation of this preliminary finding will require further Southern analysis with motif V probes. It is essential to

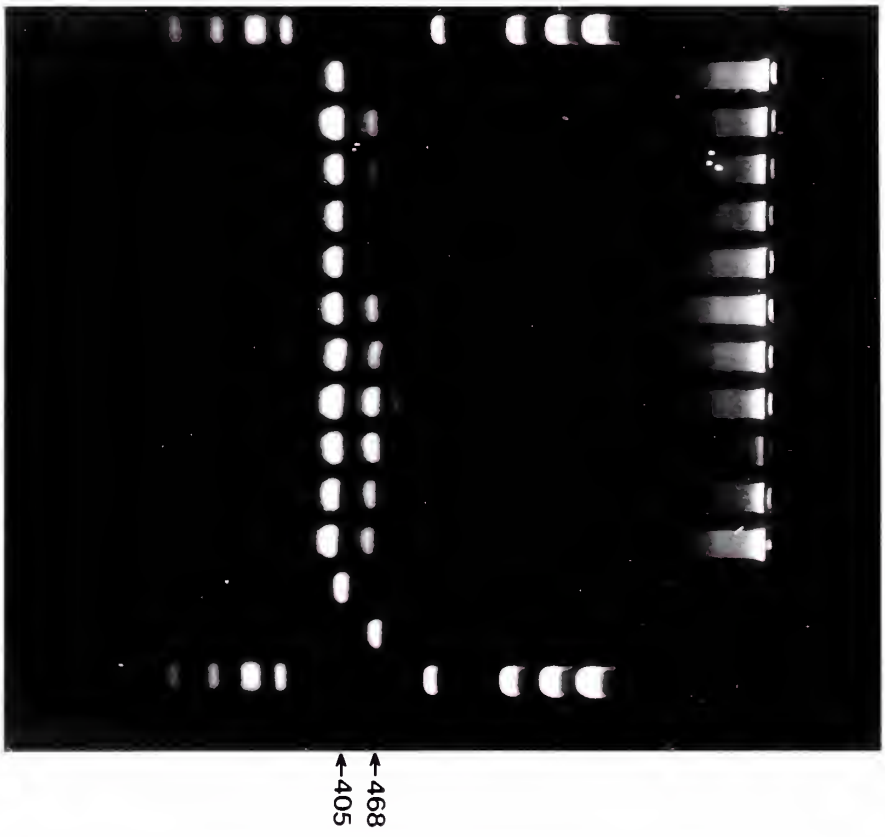
recall that the products containing motif V represent only a small fraction of the RNA (Figures 8 vs. 11). Therefore, the apparent absence of motif IV signal for the band labeled +V in Figure 13 could also be a sensitivity problem.

Although these studies require additional data to permit firm conclusions, my results strongly suggest that mRNA species capable of encoding the higher molecular weight isoform (i.e. retaining motif IV and lacking motif V, c.f. Figures 10 and 13) constitute a significant amount of the total RNA made in non-erythroid blood cells. This expression is achieved by elimination of motif V from the vast majority of species, while motif IV is retained or eliminated in roughly equal numbers of the mature mRNAs. Thus, motifs IV and V in these cells are probably not spliced in concert. (However, as discussed for Figure 13, it is possible that motif IV is selectively absent from mRNA retaining motif V.) This suggests that the predominant regulation of high vs. low molecular weight isoform production occurs via splicing of motif IV.

Expression of Motifs IV and V during Erythroid Cell Maturation. With the successful amplification of motif I in MEL cells, it seemed reasonable to expect similar results from investigation of motifs IV and V during MEL cell maturation. PCR amplification of these motifs with the same primers used in non-erythroid cells (Figure 6B) was attempted using cDNA templates generated from the mRNA of MEL cells collected at 0,

27, 48, 60, 73, 86, 98, 110, 122, 167, and 186 hours after induction. No products were seen. These results may again be attributable to inadequate homology between mouse and human sequences in the region of mRNA where motifs IV and V are located. However, abundant evidence in the PCR literature shows that successful amplification using any given primer pair depends on far more than homology alone and that other factors (GC content, secondary structure) may be responsible. This left two choices: (1) clone and sequence the mouse protein 4.1 gene so that perfectly homologous primers could be obtained, or (2) design alternative primers with different parameters in the hope that their homology to mouse sequences or their possession of dissimilar structural would yield more favorable results. I chose to pursue the latter option, because insufficient time was available to complete the former. The new primers pairs were selected on the basis of equal length, 50% GC content, identical melting temperature, absence of 3 or more G or C bases at the 3' end, no significant secondary structure, and, where possible, no runs of 3 bases. Although I have not yet tested these primers fully, early results in murine cells show promise.

Figure 7: Ethidium bromide stained agarose gels showing induction of motif I and globin during MEL cell maturation.
A. Terminal maturation of MEL cells was induced with DMSO. PCR analysis of RNA at time points during induction identifies the beginning of motif I expression. Expected sizes of products containing and lacking motif I are 468 bp and 405 bp, respectively. For unknown reasons, the mouse fragment without motif I migrates more rapidly than the comparable human fragment.
B. Similar PCR analysis of globin RNA expression shows that motif I is expressed later than globin by nearly 48 hours. (Courtesy of Dr. Lily Lou).


0°
27°
48°
60°
73°
86°
98°
110°
122°
167°
186°
MOLT 4
Retic
405
468

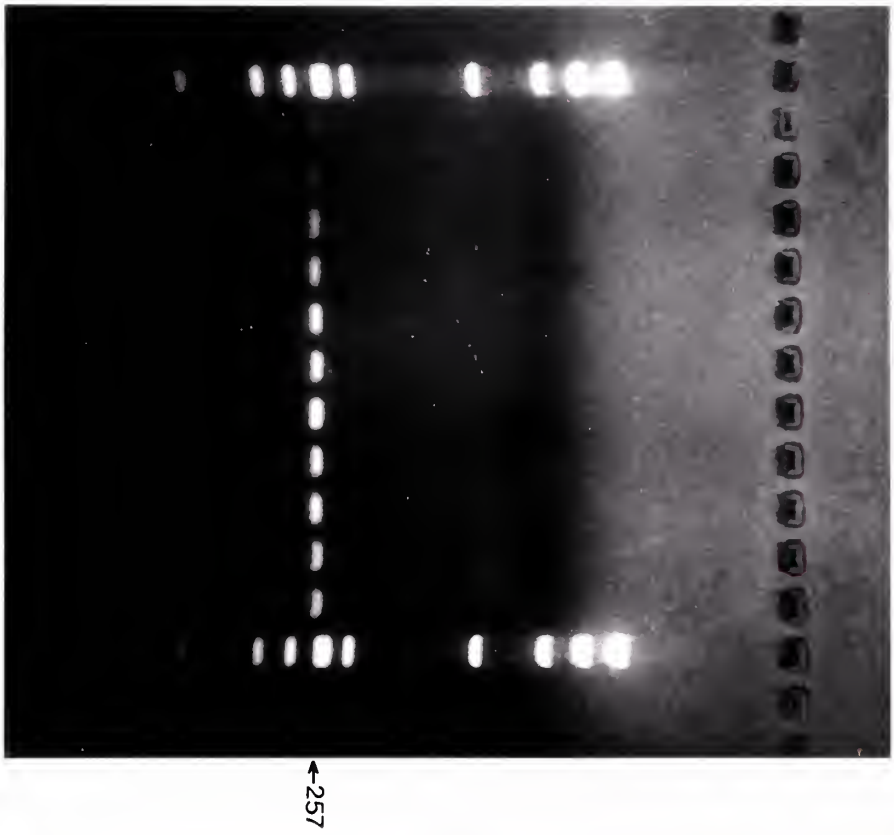

0°
27°
48°
60°
73°
86°
98°
110°
122°
167°
186°
257

Figure 8: Analysis of PCR products on ethidium bromide stained agarose gel. Expected sizes of products containing and lacking motif V are 258 bp and 178 bp, respectively. The fragments observed here are between 118 bp and 194 bp, a size consistent with absence of motif V.

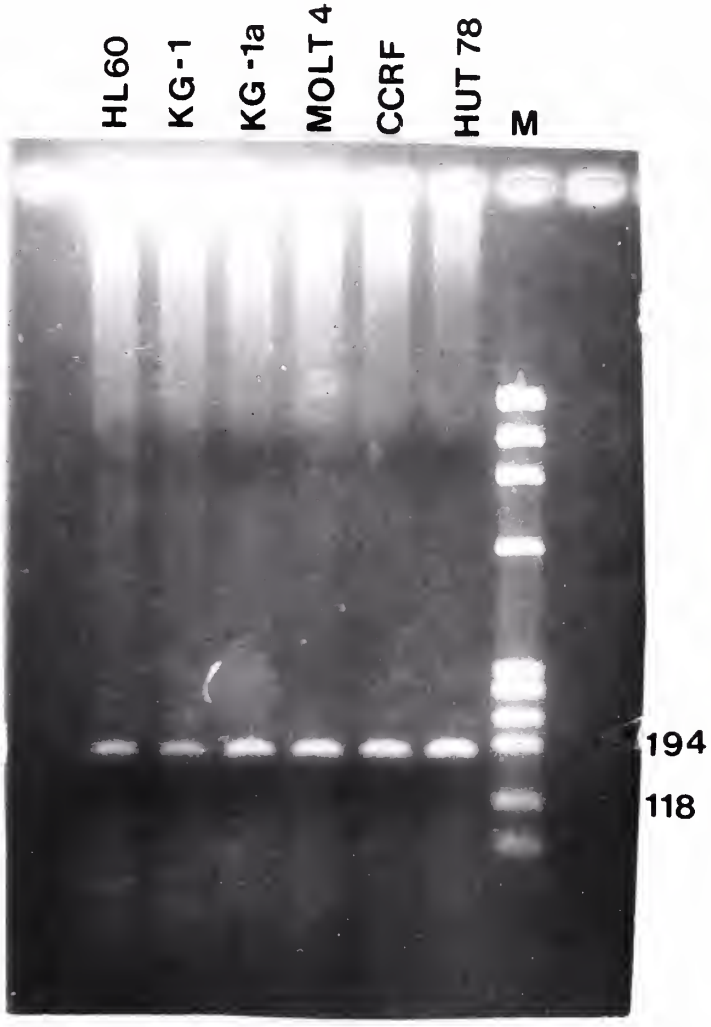


Figure 9: Analysis of PCR products on ethidium bromide stained agarose gel. Expected sizes of products containing and lacking motif IV are 96 bp and 79 bp, respectively. The fragments observed here are of two sizes: (1) between 72 bp and 118 bp and (2) roughly equal to 72 bp. These bands indicate that motif IV is present in some products and absent in others, often within a single cell line.

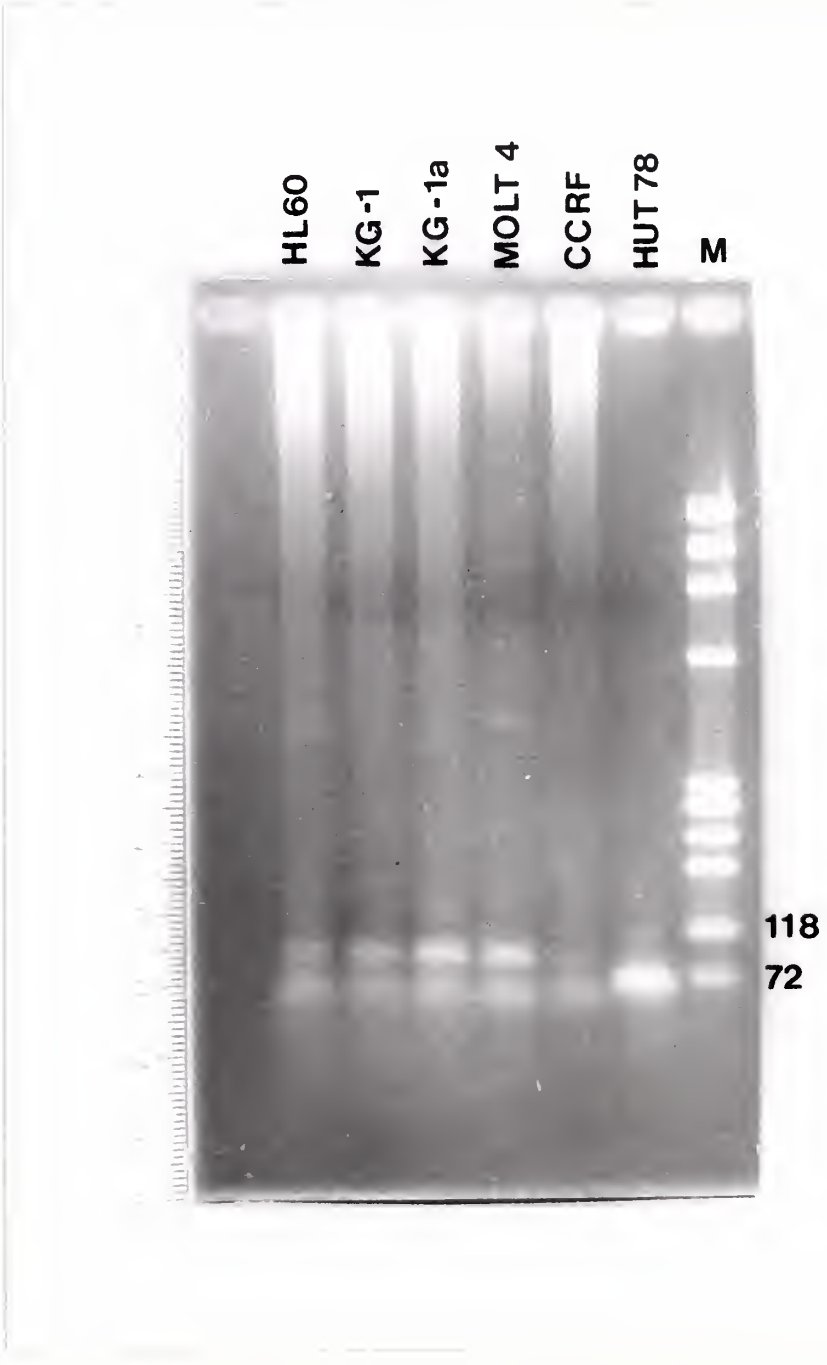


Figure 10: Analysis of PCR products on ethidium bromide stained agarose gel. Expected sizes of products containing motif V are 626 bp and 609 bp, while those lacking motif V are 546 bp and 529 bp. Since the fragments observed here all fall below the 603 bp marker, it can be concluded that they lack motif V. Presence or absence of motif IV cannot be determined from this gel, as a 17 bp difference is indistinguishable in fragments of this size.

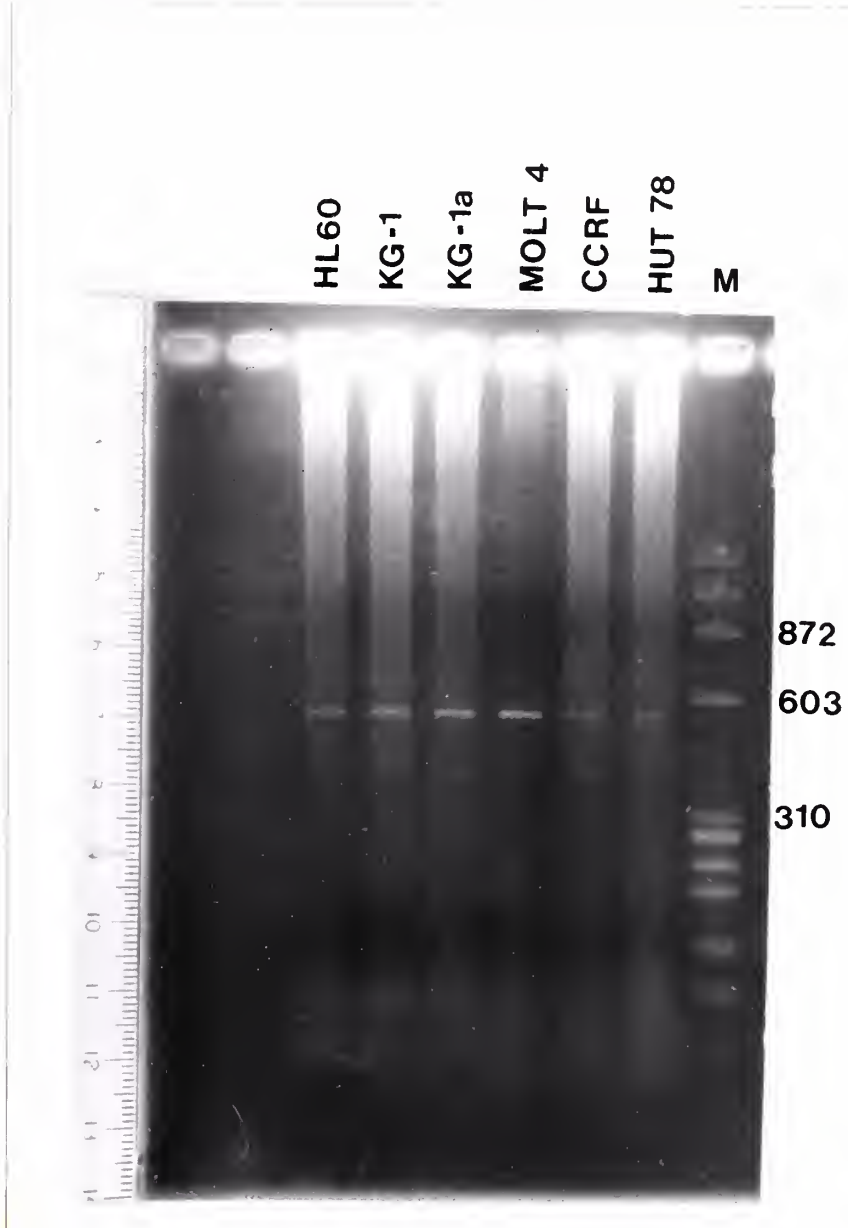


Figure 11: Southern blot analysis of PCR products. Predicted sizes of products containing and lacking motif V are indicated. The lower arrow corrsponds to the band seen on the ethidium bromide stained agarose gel. While the ethidium bromide stained agarose gel demonstrated that a majority of fragments lack motif V, hybridization with a motif V-specific probe shows that a minority of species containing motif V are also present.

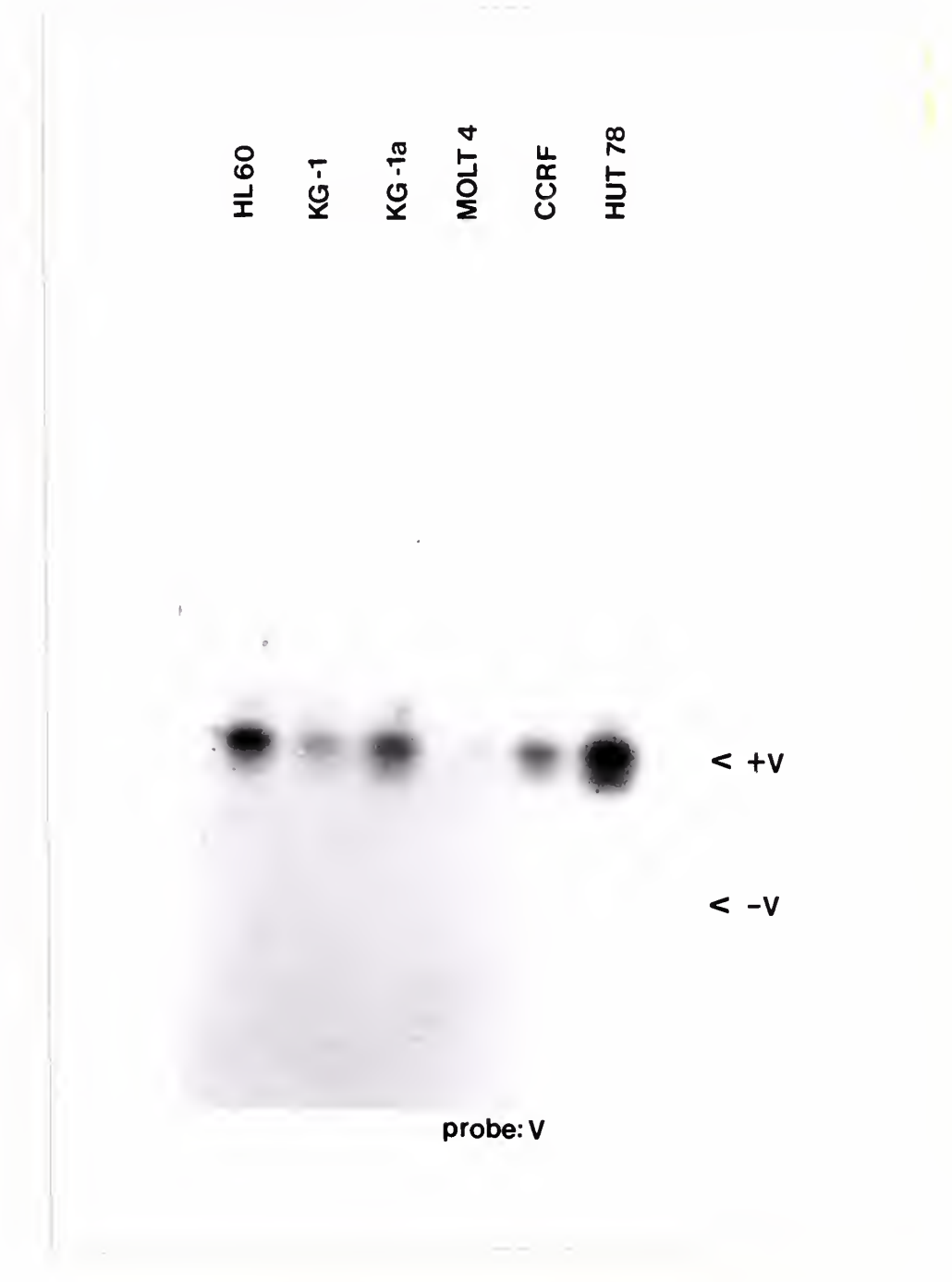


Figure 12: Southern blot analysis of PCR products. Predicted sizes of products containing and lacking motif IV are indicated. The arrows correspond to the two major bands seen on the ethidium bromide stained agarose gel. Hybridization with a motif IV-specific probe confirms that the larger band contains motif IV.

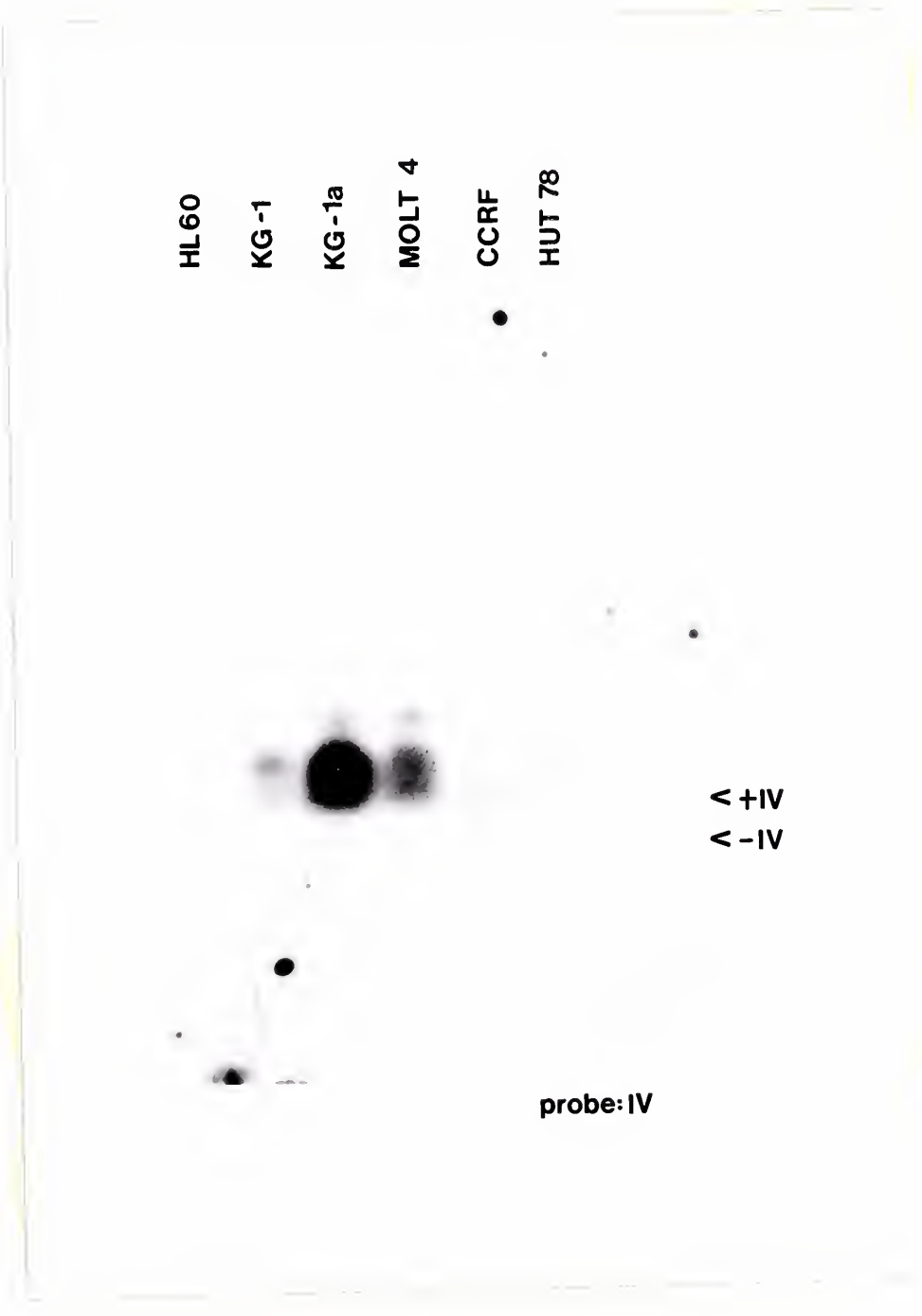
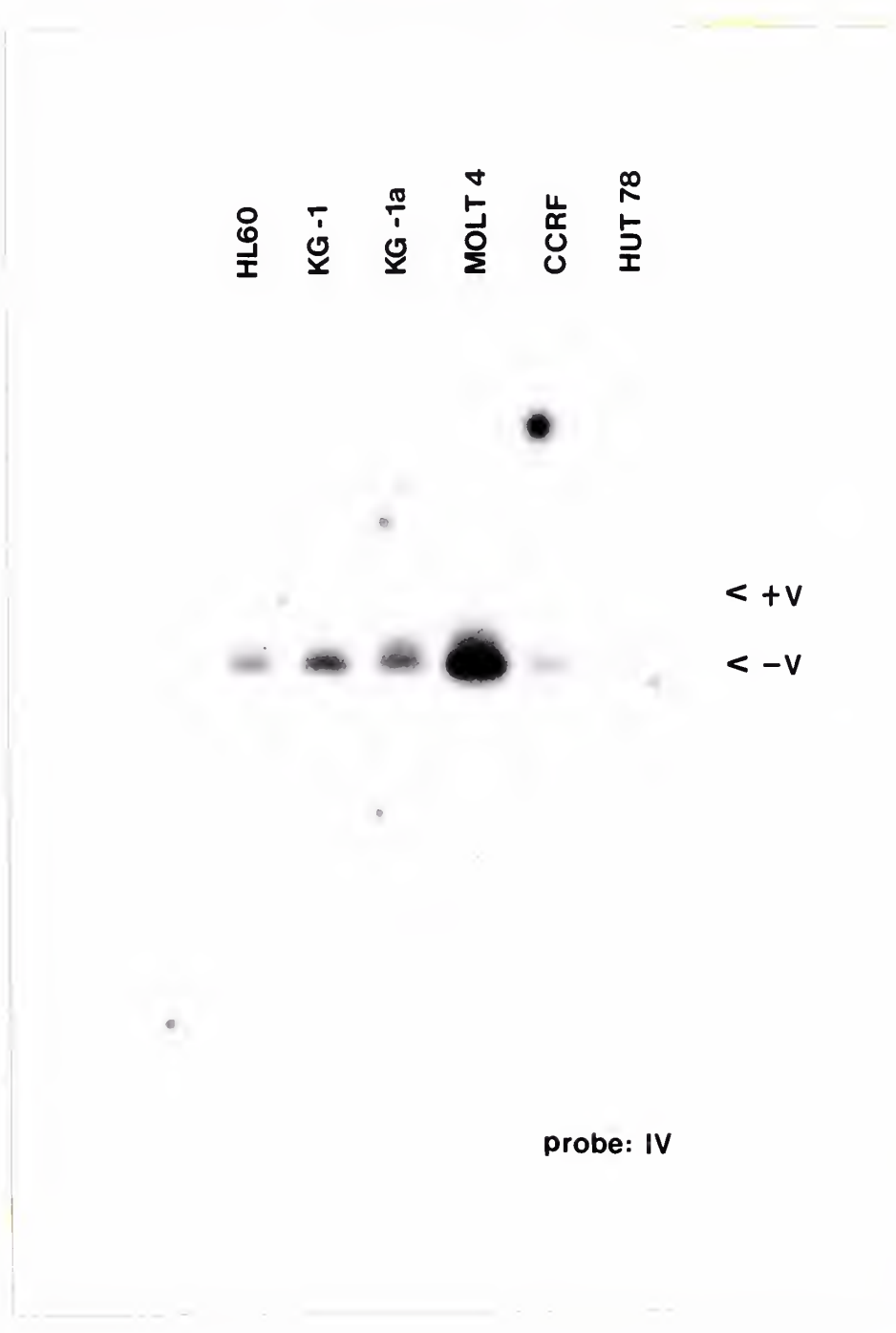


Figure 13: Southern blot analysis of PCR products. Predicted sizes of products containing and lacking motif V are indicated. Presence of motif IV is determined by hybridization to a motif IV-specific probe. On the basis of these data, it may be concluded that the products visualized on this blot contain motif IV and lack motif V.



SUMMARY and CONCLUSIONS

Protein 4.1 is a family of related cytoskeletal protein isoforms present in many cell types. These isoforms differ at specific points, designated motifs I-VIII, which are alternatively spliced nucleotide segments within protein 4.1 mRNA transcripts. Motif I is expressed in a tissue-specific manner, being found only in erythroid cells. In addition, a high molecular weight variant of protein 4.1, which localizes to nuclear as well as cytoplasmic structures of non-erythroid cells, is generated by concerted splicing of the two upstream motifs, IV and V.

The work presented in this thesis compared the pattern of mRNA splicing of motifs IV and V in erythroid and non-erythroid cells and investigated the timing of splicing events during erythroid maturation. PCR amplification of regions of mRNA spanning motifs IV and V in non-erythroid cells suggest that a substantial portion of transcripts encode the high molecular weight variant. Southern blotting revealed that transcripts for low molecular weight variants were also generated, and that their production usually occurs via motif IV deletion, rather than motif V insertion. These results provide the first indication of the mechanism by which cells regulate the production of high and low molecular weight protein 4.1 isoforms.

This research also addressed the developmental regulation of splicing events during erythroid maturation. On the basis

of my early findings from RNase protection assays and Northern blotting, another investigator employed PCR amplification to demonstrate the induction of motif I splicing between 86 and 110 hours after induction of MEL cells. Motif I induction follows that of the major erythrocyte protein, globin. Similar PCR amplification studies of motif IV and V expression during erythroid maturation have been unsuccessful.

My findings strongly imply that subsequent efforts to study the regulation of production of the 135 kD protein 4.1 isoform should be concentrated on the sequences surrounding motif IV. I have provided preliminary but highly suggestive evidence that the decision to include or exclude this motif from mature mRNA is the major determinant controlling this isoform. Therefore, this region is most likely to yield the cis-acting signals or the binding sites of trans-acting factors that regulate alternative splicing and production of this large isoform. These can be dissected by *in vitro* site-directed mutagenesis. The tissue-specificity and developmental behavior of the protein 4.1 mRNA precursor provides a unique opportunity for study of the control of alternative mRNA splicing in a multi-functional tissue-specific protein family. My results suggest that control of this family may reside in only a selected subset of the alternatively spliced motifs.

REFERENCES

Alloisio, N., Morle, L., Dorleac, E., Gentilhomme, O., Bachir, D., Guetarni, D., Colonna, P., Bost, M., Zouaoui, Z., Roda, L., Roussel, D., and Delaunay, J. (1985). The heterozygous form of 4.1(-) hereditary elliptocytosis [the 4.1(-) trait]. Blood 65,46-51.

Anderson, R.A., and Lovrien, R.E. (1984). Glycophorin is linked by band 4.1 protein to the human erythrocyte membrane skeleton. Nature 307,655-658.

Anderson, R.A., and Marchesi, V.T.. (1985). Regulation of the association of membrane skeletal protein 4.1 with glycophorin by a polyphosphoinositide. Nature 318,295-298.

Aster, J.C., Brewer, G.J., and Maisel, H. (1986). The 4.1-like proteins of the bovine lens: spectrin-binding proteins closely related in structure to red blood cell protein 4.1. The Journal of Cell Biology 103,115-122.

Baines, A.J., and Bennett, V. (1986). Synapsin I is a microtubule-bundling protein. Nature 319,145-147.

Baines, A.J., and Bennett, V. (1985). Synapsin I is a spectrin-binding protein immunologically related to erythrocyte protein 4.1. Nature 315,410-413.

Chang, H., Langer, P.J., and Lodish, H.F. (1976). Asynchronous synthesis of erythrocyte membrane proteins. Proceedings of the National Academy of Sciences 73,3206-3210.

Cohen, C.M., Foley, S.F., and Korsgren, C. (1982). A protein immunologically related to erythrocyte band 4.1 is found on stress fibres of non-erythroid cells. Nature 299,648-650.

Conboy, J.G., Chan, J., Mohandas, N., and Kan, Y.W. (1988). Multiple protein 4.1 isoforms produced by alternative splicing in human erythroid cells. Proceedings of the National Academy of Sciences 85,9062-9065.

Conboy, J., Kan, Y.W., Shohet, S.B., and Mohandas, N. (1986a). Molecular cloning of protein 4.1, a major structural element of the human erythrocyte membrane skeleton. Proceedings of the National Academy of Sciences 83,9512-9516.

Conboy, J., Marchesi, S., Kim, R., Agre, P., Kan, Y.W., and Mohandas, N. (1990). Molecular analysis of insertion/deletion mutations in protein 4.1 in elliptocytosis: II. determination of molecular genetic

origins of rearrangements. Journal of Clinical Investigation 86,524-530.

Conboy, J., Mohandas, N., Tchernia, G., and Kan, Y.W. (1986b). Molecular basis of hereditary elliptocytosis due to protein 4.1 deficiency. The New England Journal of Medicine 315,680-685.

Correas, I., Leto, T.L., Speicher, D.W., and Marchesi, V.T. (1986). Identification of the functional site of erythrocyte protein 4.1 involved in spectrin-actin associations. The Journal of Biological Chemistry 261,3310-3315.

Davies, G.E, and Cohen, C.M. (1985). Platelets contain proteins immunologically related to red cell spectrin and protein 4.1. Blood 65,52-59.

Davis, L.G., Dibner, M.D., and Battey, J.F. (1986). Basic methods in molecular biology. (Elsevier Science Publishing Co., Inc., New York, NY).

Fairbanks, G., Steck, T.L., and Wallach, D.F.H. (1971). Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry 10,2606-2617.

Goodman, S.R., Casoria, L.A., Coleman, D.B., and Zagon, I.S. (1984). Identification and location of brain protein 4.1. Science 224,1433-1436.

Goodman, S.R., Yu, J., Whitfield, C.F., Culp, E.N., and Posnak, E.J. (1982). Erythrocyte membrane skeletal protein bands 4.1 a and b are sequence-related phosphoproteins. The Journal of Biological Chemistry 257,4564-4569.

Granger, B.L., and Lazarides, E. (1985). Appearance of new variants of membrane skeletal protein 4.1 during terminal differentiation of avian erythroid and lenticular cells. Nature 313,238-241.

Granger, B.L., and Lazarides, E. (1984). Membrane skeletal protein 4.1 of avian erythrocytes is composed of multiple variants that exhibit tissue-specific expression. Cell 37,595-607.

Horne, W.C., Tang, T.K., Marchesi, V.T., and Benz, E.J., Jr. Selective expression of an exon alters the spectrin-binding properties of protein 4.1, in press.

Leto, T.L., and Marchesi, V.T. (1984). A structural model of human erythrocyte protein 4.1. The Journal of Biological Chemistry 259,4603-4608.

Leto, T.L., Pratt, B.M., and Madri, J.A. (1986). Mechanisms of cytoskeletal regulation: modulation of aortic endothelial cell protein band 4.1 by the extracellular matrix. Journal of Cellular Physiology 127,423-431.

Liu, S.-C., Derick, L.H., and Palek, J. (1987). Visualization of the hexagonal lattice in the erythrocyte membrane skeleton. The Journal of Cell Biology 104,527-536.

Marchesi, S.L., Conboy, J., Agre, P., Letsinger, J.T., Marchesi, V.T., Speicher, D.W., and Mohandas, N. (1990). Molecular analysis of insertion/deletion mutations in protein 4.1 in elliptocytosis: I. biochemical identification of rearrangements in the spectrin/actin binding domain and functional characterizations. Journal of Clinical Investigation 86,516-523.

Marks, P.A., and Rifkind, R.A. (1978). Erythroleukemia differentiation. Annual Review of Biochemistry 47,419-448.

Ngai, J., Stack, J.H., Moon, R.T., and Lazarides, E. (1987). Regulated expression of multiple chicken erythroid membrane skeletal protein 4.1 variants is governed by differential RNA processing and translational control. Proceedings of the National Academy of Sciences 84,4432-4436.

Pasternack, G.R., Anderson, R.A., Leto, T.L., and Marchesi, V.T. (1985). Interactions between protein 4.1 and band 3: an alternative binding site for an element of the membrane skeleton. The Journal of Biological Chemistry 260,3676-3683.

Spiegel, J.E., Beardsley, D.S., Southwick, F.S., and Lux, S.E. (1984). An analogue of the erythroid membrane skeletal protein 4.1 in nonerythroid cells. The Journal of Cell Biology 99,886-893.

Steck, T.L. (1974). The organization of proteins in the human red blood cell membrane: a review. The Journal of Cell Biology 62,1-19.

Takakuwa, Y., Tchernia, G., Rossi, M., Benabadji, M., and Mohandas, N. (1986). Restoration of normal membrane stability to unstable protein 4.1-deficient erythrocyte membranes by incorporation of purified protein 4.1. Journal of Clinical Investigation 78,80-85.

Tang, T.K. (1988a). Molecular cloning and tissue specific regulation of erythroid and non-erythroid membrane skeletal protein 4.1. Doctoral Thesis. Department of Human Genetics, Yale University, New Haven, Conn.

Tang, T.K., Benz, E.J., Jr., and Marchesi, V.T. Identification of a protein 4.1 isoform in human lymphocytes with a nuclear localization, in press.

Tang, T.K., Leto, T.L., Correas, I., Alonso, M.A., Marchesi, V.T., and Benz, E.J., Jr. (1988b). Selective expression of an erythroid-specific isoform of protein 4.1. Proceedings of the National Academy of Sciences 85,3713-3717.

Tang, T.K., Qin, Z., Leto, T., Marchesi, V.T., and Benz, E.J., Jr. (1990). Heterogeneity of mRNA and protein products arising from the protein 4.1 gene in erythroid and nonerythroid tissues. The Journal of Cell Biology 110,617-624.

Tchernia, G., Mohandas, N., and Shohet, S.B. (1981). Deficiency of skeletal membrane protein band 4.1 in homozygous hereditary elliptocytosis: implications for erythrocyte membrane stability. The Journal of Clinical Investigation 68,454-460.

Ungewickell, E., Bennett, P.M., Calvert, R., Ohanian, V., and Gratzer, W.B. (1979). *In vitro* formation of a complex between cytoskeletal proteins of the human erythrocyte. Nature 280,811-814.

Yew, N.S., Choi, H.-R., Gallarda, J.L., and Engel, J.D. (1987). Expression of cytoskeletal protein 4.1 during avian erythroid cellular maturation. Proceedings of the National Academy of Sciences 84,1035-1039.

HARVEY CUSHING / JOHN HAY WHITNEY
MEDICAL LIBRARY

MANUSCRIPT THESES

Unpublished theses submitted for the Master's and Doctor's degrees and deposited in the Medical Library are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but passages must not be copied without permission of the authors, and without proper credit being given in subsequent written or published work.

This thesis by has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

NAME AND ADDRESS DATE